Quelques propriétés associées à la suite de Prouhet-Thue-Morse

Jean Moussa(*)

Cet article a été motivé par la résolution de l'exercice suivant :

La suite des fractions :

$$\frac{1}{2}\ ;\ \frac{\left(\frac{1}{2}\right)}{\left(\frac{3}{4}\right)};\frac{\left(\left({}^{1}\!/_{2}\right)\Big/\left({}^{3}\!/_{4}\right)\right)}{\left(\left({}^{3}\!/_{4}\right)\Big/\left({}^{7}\!/_{8}\right)\right)}\ ;\ \text{etc}$$

admet- t-elle une limite ? et, si oui, laquelle ?

La question avait été imprudemment proposée sous la rubrique « De-ci de-là » (BV nº 498-4) comme exercice à destination de nos élèves..

Résoudre cet exercice nécessite d'aborder l'étude de la suite de Prouhet-Thue-Morse. André Mery, mathématicien amateur averti, nous a envoyé une solution que l'on peut lire sur (référence du site).

L'article ci-dessous propose une familiarisation avec la suite de Thue et une de ses propriétés remarquables (parties I et II). Dans la troisième partie, nous suivons le schéma de démonstration suivi par André Méry, mais sous une présentation différente.

Nous présentons enfin deux algorithmes de construction de la suite de Thue.

I : La suite de Thue.

Nous définissons une suite, appelée « suite de Thue » de la manière suivante :

Pour tout entier naturel k, si l'écriture en base 2 de k est : $k = \varepsilon_p\varepsilon_{p-1}\varepsilon_{p-2}\cdots\varepsilon_1\varepsilon_0$, avec $\varepsilon_i = 0$ ou 1, alors nous définissons le terme correspondant de la suite de Thue par : $t_k = \sum\varepsilon_i \bmod(2)$: autrement dit, le k-ième terme de la suite de Thue est le reste modulo 2 de la somme des chiffres dans l'écriture binaire de k.
On peut aussi présenter cette suite sous une autre forme en prenant $v_k = 1 - 2t_k$, ce qui représente la parité ou l'imparité par les nombres ± 1.

(*) jean5moussa@gmail.com

k	0	1	2	3	4	5	6	7
en base 2	0	1	10	11	100	101	110	111
t_k	0	1	1	0	1	0	0	1
v_k	1	−1	−1	1	−1	1	1	−1

Il en découle assez rapidement trois propriétés remarquables :

Première propriété. Pour tout naturel k, on a : $t_{2k} = t_k$ (sous l'autre forme, on a aussi $v_{2k} = v_k$). En effet l'écriture binaire de $2k$ est identique à celle de k au zéro final près, de sorte que la somme des chiffres reste inchangée, ainsi que sa parité.

Deuxième propriété. Pour tout naturel k, on a : $t_{2k+1} = 1 - t_{2k}$ (sous l'autre forme cela devient $v_{2k+1} = -v_{2k}$) c'est-à-dire que les deux termes consécutifs en question sont ou bien (01) ou bien (10). En effet, l'écriture binaire de $2k + 1$ est identique à celle de $2k$ au dernier chiffre près, qui vaut toujours 0 dans $2k$, et 1 dans $2k + 1$ de sorte que la somme des chiffres augmente toujours de 1 en passant de l'un à l'autre ; sa parité ne peut donc que changer.

On peut étendre cette propriété en considérant les quatre termes successifs t_{4k}, t_{4k+1}, t_{4k+2}, t_{4k+3}, et les deux derniers chiffres des indices correspondants, qui, en base 2, sont toujours {00, 01, 10, 11}. On trouve ainsi :

$$\left(v_{4k}, v_{4k+1}, v_{4k+2}, v_{4k+3}\right) = \left(1,-1,-1,1\right) \text{ ou bien } \left(-1,1,1,-1\right)$$
$$\left(t_{4k}, t_{4k+1}, t_{4k+2}, t_{4k+3}\right) = \left(1,0,0,1\right) \text{ ou bien } \left(0,1,1,0\right)$$

Troisième propriété. À tout entier k de l'intervalle $\left[0 \cdots 2^p\right[$ on fait correspondre un entier de l'intervalle $\left[2^p \cdots 2^{p+1}\right[$ par : $0 \le k < 2^p \Leftrightarrow 2^p \le k + 2^p < 2^{p+1}$.

Si $k = \varepsilon_p \varepsilon_{p-1} \varepsilon_{p-2} \cdots \varepsilon_1 \varepsilon_0$ alors $k + 2^p = 1\varepsilon_p \varepsilon_{p-1} \varepsilon_{p-2} \cdots \varepsilon_1 \varepsilon_0$ et le « 1 » change la parité de la somme des chiffres. On en conclut que, dans la suite de Thue, les deux paquets des termes correspondants sont « symétriques », les deux valeurs possibles étant échangées. Ceci met en évidence un procédé de construction de la suite, « par blocs » successifs de taille doublée à chaque étape :

(0)(1)(10)(1001)(1001 0110)(1001 0110 0110 1001) …

Cette troisième propriété est celle qui fait apparaître la suite de Thue dans la construction de la fraction hiérarchisée de la partie III.

On peut remarquer ici la forme particulière des groupements de quatre termes consécutifs de la suite, qui illustre l'extension donnée ci-dessus à la deuxième propriété.

II Une propriété remarquable de la suite de Thue.

La suite étudiée ici a été définie par Axel Thue dans un article publié en 1912, puis redécouverte par Marston Morse en 1921 (sans rapport avec Samuel Morse, le père de l’alphabet du même nom). Elle était déjà apparue en 1851 dans les recherches d’Eugène Prouhet à propos d’un problème de théorie des nombres, d’où son nom de « suite de Prouhet-Thue-Morse ».

La combinatoire des mots est un sujet ouvert au début du XX^{e} siècle, et dont l’étude actuelle est particulièrement intense, car elle est utilisée dans tous les logiciels automatiques d’analyse, de reconnaissance de texte et de traitement du langage.

On s’y intéresse à des messages, c'est-à-dire des successions de signes, ce qui constitue un modèle mathématique évident de tout texte écrit, tel celui que vous êtes en train de lire. L’espace est ici un signe parmi d’autres. Les signes sont pris dans un alphabet qui peut être celui que nous utilisons en français, mais qui peut aussi être défini de manière arbitraire. L’alphabet réduit aux deux signes 0 et 1 présente l’avantage d’être le plus simple, et il joue un rôle essentiel depuis les développements de l’informatique. On parlera de « mots » - ce sont nos mots usuels – et de « motifs », qui sont des assemblages de signes et qu’on rencontre comme parties d’un message. Les problèmes abordés sont très divers, depuis des problèmes de dénombrement jusqu’aux tests optimaux pour chercher ou comparer des séquences.

Soit un alphabet comportant deux lettres, disons 0 et 1. Est-il possible d’écrire un message aussi long que l’on voudra et qui ne comporte aucun motif répété deux fois à la suite ? La réponse est non, et on comprend aisément pourquoi :

Si l’on essaie en commençant par 0, on ne doit pas continuer par 0, sinon cela fait (0)(0), donc il faut poursuivre par1 : 01. Puis on ne doit pas continuer par 1, sinon cela ferait 0(1)(1), donc la seule solution est 010 ; ensuite 0100 fait apparaître 01(0)(0), tandis que 0101 = (01)(01) : on échoue donc toujours à construire une telle suite à partir de seulement quatre signes.

Thue s’est posé la question de l’existence de messages ne comportant aucune triple répétition ; par exemple, le message 0110(011)(011)(011)10 comporte une telle triple répétition. Thue a démontré que sa suite était un exemple d’existence.

Nous allons démontrer une propriété un peu plus forte : la suite de Thue ne contient aucun motif de la forme {0M0M0} ou de la forme {1M1M1}, où M désigne un motif de longueur quelconque. Ceci interdit bien sûr toute triple répétition d’un motif, car ce motif commencerait par 0 ou 1, et en ce cas il y aurait donc dans la suite de Thue un motif de la forme {0M0M0M} ou {1M1M1M}, ce qui vient d’être exclu. Tous les raisonnements qui suivent fonctionneront de la même manière pour {1M1M1} et pour {0M0M0}, ce qui dispense d’étudier séparément les deux cas.

(A) Supposons l’existence d’un motif {0M0M0} où M = $m_1m_2…m_2$ est de longueur paire, $2p$, $p > 0$. L’indice du premier « 0 » est soit pair soit impair.

(A1) **Dans le cas où cet indice est pair, $2k$**, nous avons le tableau suivant :

2k	2k+1	…..	2k+2p	2k+2p+1	2k+2p+2	2k+2p+3	…..	2k+4p	2k+4p+1	2k+4p+2
0	m_1	…..	m_{2p}	0	m_1	m_2	…..	m_{2p-1}	m_{2p}	0
---------(a)			--------- (*)		---------(b)			---------(y)		

où l’on a signalé sur la ligne du bas les « couples » formés par un indice pair et l'indice impair consécutif ; par l'application à tous ces couples de la propriété $t_{2k+1} = 1 - t_{2k}$ il apparaît une impossibilité, puisque l'on trouve ainsi par (a) $m_1 = 1$, puis par (b) $m_2 = 0$, etc.. en zigzaguant jusqu'à (y) qui donne $m_{2p} = 0$, qui contredit enfin (*).

(A2) Dans le cas où l'indice du premier « 0 » est impair, on remarque que l'indice du dernier « 0 » sera impair. On va donc construire le même tableau, dans lequel les couples de la ligne du bas seront simplement décalés :

2k-1	2k	2k+1	…	2k+2p-1	2k+2p	2k+2p+1	2k+2p+2	2k+2p+3	…	2k+4p	2k+4p+1
0	m_1	m_2	…	m_{2p}	0	m_1	m_2	m_3	…	m_{2p}	0
	---------				---------		---------			---------	

et, en utilisant de la même manière la propriété $t_{2k+1} = 1 - t_{2k}$, on retrouve la même impossibilité.

(B) Supposons maintenant l'existence d'un motif {0M0M0} où M est de longueur impaire, et soi $2p + 1$ cette longueur. On suppose aussi que $2p + 1$ est le plus petit entier impair possible pour qu'un tel motif apparaisse. L'indice du premier « 0 » est à nouveau soit pair soit impair.

(B1) Dans le cas où cet indice est pair, si $p = 0$, nous sommes en présence du motif {01010}, avec trois zéros sur des indices pairs, ce qui est impossible car d'après la propriété $t_{2k} = t_k$, cela ferait apparaître trois zéros consécutifs plus tôt dans la suite, ce qui est exclu par la propriété $t_{2k+1} = 1 - t_{2k}$. Le cas {00000} est lui aussi exclu. On considère maintenant $p \geq 1$.

Du tableau :

2k	2k+1	…	2k+2p+1	2k+2p+2	2k+2p+3	…	2k+4p+3	2k+4p+4
0	m_1	…	m_{2p+1}	0	m_1	…	m_{2p+1}	0

on extrait les colonnes des indices pairs :

2k	2k+2	…	2k+2p	2k+2p+2	2k+2p+4	…	2k+4p+2	2k+4p+4
0	m_2	…	m_{2p}	0	m_2	…	m_{2p}	0

Ce qui donne, en utilisant la propriété $t_{2k} = t_k$, un motif {0M0M0} où M est maintenant de longueur p. On a vu que p ne peut pas être pair strictement positif, et comme on l'a choisi pour que $2p + 1$ soit le plus petit entier impair pour lequel un motif se présente, il y a contradiction.

Le raisonnement fonctionne de la même manière avec un motif {1M1M1}.

(B2) Dans le cas où l'indice de départ est impair, le terme précédant le premier « zéro » du motif est d'indice pair et, en utilisant la relation $t_{2k+1} = 1 - t_{2k}$, on voit qu'il vaut nécessairement 1, et il en va de même pour m_{2p+1} dans le tableau suivant :

2k-2	2k-1	2k	…	2k+2p	2k+2p+1	2k+2p+2	…	2k+4p+2	2k+4p+3
1	0	m_1	…	$m_{2p+1} = 1$	0	m_1	…	$m_{2p+1} = 1$	0

Ce tableau fait alors apparaître un motif du type {1N1N1}, où N = $\{0 m_1 m_2 \ldots m_{2p}\}$ est encore de longueur $2p + 1$, mais le motif débute à un indice pair, ce qui, on vient de le voir, est impossible !

III Suite de fractions hiérarchisées.

Nous revenons à la suite de fractions présentée au début de cet article.
On peut toujours écrire ces fractions sous une forme « simple », la troisième étant par exemple égale à $\frac{1\cdot 4\cdot 6\cdot 7}{2\cdot 3\cdot 5\cdot 8}$. La n-ième fraction fait ainsi apparaître les entiers de 1 à 2^n, tantôt au numérateur tantôt au dénominateur. La n+1-ième s'obtient en lui accolant une « image » de la n-ième, inversée (car placée en dénominateur), et avec les mêmes entiers augmentés de 2^n. Cette règle reproduit donc très exactement la méthode de construction de la suite de Thue par blocs successifs de taille doublée à chaque étape.

La n-ième fraction de cette suite est donc : $u_n = \prod_{k=0}^{2^n-1}(1+k)^{v_k}$.

Définissons une suite de fonctions f_n, $n \in \mathbb{N}$ par: $f_n(x) = \prod_{k=2}^{n-1}(x+k)^{v_k}$ (nous commençons à l'indice 2 pour des raisons ultérieures de commodité).

On voit que $u_n = 1^{v_0}2^{v_1} f_{2^n}(1) = \frac{1}{2} f_{2^n}(1)$.

La question consistant à étudier la convergence de $\{u_n, n \in \mathbb{N}\}$, on va s'intéresser à celle de la suite de fonctions.

Soit $g_n(x) = \ln(f_n(x)) = \sum_{k=2}^{n-1} v_k \ln(x+k)$. L'alternance des signes dans la suite v_k suggère d'étudier les sommes de quatre termes successifs séparant les indices multiples de 4.

On a vu (extension de la troisième propriété) que les quatre termes v_* entre deux multiples de 4 ne peuvent prendre que deux formes : {1, −1, −1, 1}, ou {−1, 1, 1, −1}. On a donc :

$$g_{4(p+1)} - g_{4p}(x) = v_{4p}\left[\ln(x+4p) - \ln(x+4p+1) - \ln(x+4p+2) + \ln(x+4p+3)\right]$$
$$= v_{4p} \ln\left[\frac{(x+4p)(x+4p+3)}{(x+4p+1)(x+4p+2)}\right] = v_{4p} \ln\left[1 - \frac{2}{(x+4p+1)(x+4p+2)}\right]$$

Ce qui, lorsque $p \to \infty$, est un infiniment petit, et, en valeur absolue, équivaut à $\frac{1}{18p^2}$.

Par comparaison avec la série de Riemann d'exposant 2, on voit que, pour tout $x \in [0,+\infty[$, la série de terme général $g_{4(p+1)}(x) - g_{4p}(x)$ est absolument convergente.

Cela implique que la suite des fonctions $g_{4p}, p \in \mathbb{N}$ converge vers une fonction définie sur l'intervalle $[0,+\infty[$.

Revenant à $f_n \exp(g_n)$, on en déduit que la suite des fonctions f_{4p} converge sur l'intervalle $[0,+\infty[$ vers une fonction strictement positive F. Par conséquent, la suite u_n, qui est extraite de $\frac{1}{2} f_{4p}(1)$, converge vers un nombre strictement positif, $\frac{1}{2}$ F(1).

On peut étendre ce résultat à la suite des fonctions g_{2p}, et donc aussi à f_{2p}. En effet, il y a seulement à rajouter les termes d'indice pair non multiples de 4, donc les $4p + 2$; mais ceux-ci sont « éloignés » des termes d'indice $4p$ par des sommes de la forme $\pm\left(\ln(x+4p)-\ln(x+4p+)\right)$, grâce à l'alternance des signes de la suite v. Or cet ajout tend vers zéro lorsque $p \to +\infty$. On peut donc aussi écrire :

$$\mathrm{F}(x) = \lim_{p\to+\infty} f_{2p(x)}.$$

$$\mathrm{F}(1) = \lim_{p\to+\infty} \prod_{k=2}^{2p-1} (1+k)^{v_k}.$$

La suite $f_{2p}(0) = \prod_{k=2}^{2p-1} k^{v_k}$ converge vers le nombre strictement positif F(0).

Considérons maintenant le produit :

$$f_{4p}(0) f_{4p}(1) = \left[2^{v_2}3^{v_3}4^{v_4}\cdots(4p-1)^{v_{4p-1}}\right]\left[3^{v_2}4^{v_3}\cdots(4p)^{v_{4p-1}}\right].$$

On le réordonne :

$$f_{4p}(0) f_{4p}(1) = 2^{v_2}3^{v_3+v_2}4^{v_4+v_3}\cdots(4p-1)^{v_{4p-1}+v_{4p-2}}(4p)^{v_{4p-1}}.$$

D'après la relation $v_{2k+1} = -v_{2k}$, la moitié des termes disparaissent et il reste :

$$f_{4p}(0) f_{4p}(1) = 2^{v_2}4^{v_4+v_3}6^{v_6+v_5}8^{v_8+v_7}\cdots(4p-2)^{v_{4p-2}+v_{4p-3}}(4p)^{v_{4p-1}}.$$

On réordonne à nouveau, toujours pour utiliser la relation $v_{2k+1} = -v_{2k}$, mais différemment :

$$= \left(2^{v_2}4^{v_3}\right)\left(4^{v_4}6^{v_5}\right)\left(6^{v_6}8^{v_7}\right)\cdots\left((4p-4)^{v_{4p-4}}(4p-2)^{v_{4p-3}}\right)\left((4p-2)^{v_{4p-2}}(4p)^{v_{4p-1}}\right)$$

$$= \left(2^{v_2}4^{-v_2}\right)\left(4^{v_4}6^{-v_4}\right)\left(6^{v_6}8^{-v_8}\right)\cdots\left((4p-2)^{v_{4p-2}}(4p)^{-v_{4p-2}}\right).$$

… expression où l'on remarque que les parenthèses de la forme $\left(2^{v_2}4^{-v_2}\right)$ etc.. se simplifient, en tant que fractions et par la relation $v_{2k} = v_k$, ce qui donne :

$$f_{4p}(0) f_{4p}(1) = \left(1^{v_1}2^{-v_1}\right)\left(2^{v_2}3^{-v_2}\right)\left(3^{v_3}4^{-v_2}\right)\cdots\left((2p-1)^{v_{2p-1}}(2p)^{-v_{2p-1}}\right).$$

D'où en réordonnant encore une fois :

$$f_{4p}(0) f_{4p}(1) = \left(1^{v_1}2^{v_2}3^{v_3}\cdots(2p-1)^{v_{2p-1}}\right)\cdot\left(2^{-v_1}3^{-v_2}4^{-v_3}\cdots(2p)^{-v_{2p-1}}\right)$$

où l'on reconnaît :

$$f_{4p}(0)f_{4p}(1)=2^{-v_1}f_{2p}(0)\left(f_{2p}(1)\right)^{-1}.$$

Comme $v_1 = -1$, et comme les suites figurant ici convergent, on en déduit :

$$\mathrm{F}(0)\mathrm{F}(1)=2\mathrm{F}(0)\left(\mathrm{F}(1)\right)^{-1},$$

mais F(0) est non nul ; donc $\mathrm{F}(1)=\sqrt{2}$, , et finalement :

$$\lim_{n\to+\infty} u_n = \frac{1}{2}\mathrm{F}(1)=\frac{1}{\sqrt{2}}.$$

Vitesse de convergence

La démonstration proposée ne fournit qu'une majoration de l'écart à la limite. Nous avons en effet majoré à l'aide d'une série de Riemann. Or le reste à l'ordre n d'une série de Riemann en $\frac{1}{n^2}$, décroît comme $\frac{1}{n}$, et donc, on peut déduire de cette démonstration que :

$$\left|u_n-\frac{1}{2}\mathrm{F}(1)\right|=\frac{1}{2}\left|f_{2^{n\le\frac{\le}{}}}(1)-\frac{1}{2}\mathrm{F}(1)\right|\le\frac{\mathrm{A}}{2^n},$$

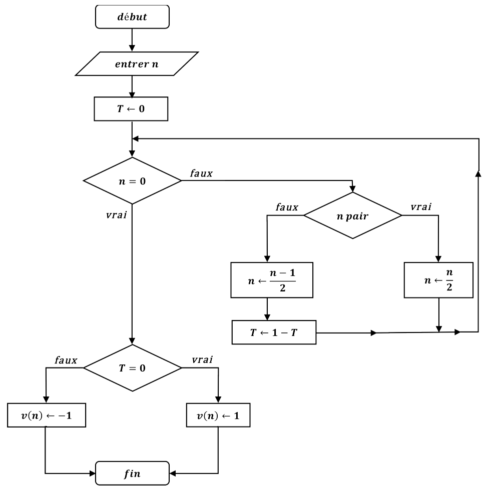
pour un certain A. L'étude numérique de la suite u_n montre que la convergence est en fait beaucoup plus rapide :

n	u_n	u_n-*lim*	$1/2^n$
6	0,707102124480..	$\approx 4{,}6.10^{-6}$	$\approx 1{,}6.10^{-2}$
7	0,707106622058..	$\approx 1{,}6.10^{-7}$	$\approx 8.10^{-3}$
8	0,707106777975..	$\approx 3{,}2.10^{-9}$	$\approx 4.10^{-3}$
9	0,707106781149..	$\approx 3{,}7.10^{-11}$	$\approx 2.10^{-3}$
10	0,707106781186..	$\approx 2{,}5.10^{-13}$	$\approx 1.10^{-3}$

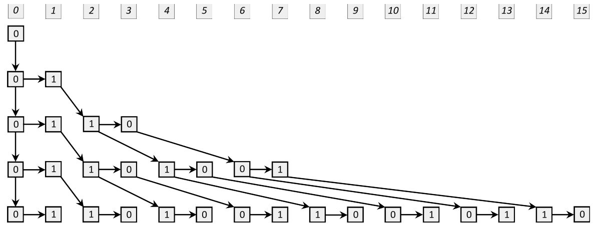
Lorsque l'on passe d'un terme au terme suivant, l'écart à la limite décroît de plus en plus vite par rapport à $\frac{1}{2^n}$. La majoration ci-dessus est faible car elle a été établie comme s'il n'y avait pas de « compensation » entre les groupements de quatre termes consécutifs, c'est-à-dire comme si la suite de Thue était périodique : (1001)(1001)(1001).

IV Algorithmes de construction de la suite de Thue

Le premier algorithme donne un terme (t ou v) de la suite dont l'ordre (n) aura été fixé à l'avance. Il fonctionne comme celui qui donne la décomposition d'un nombre en base 2, par divisions successives par deux et en collationnant les restes. Dans le sens descendant, il est donc accolé à la définition donnée dans cet article.



Mais il est intéressant de voir en œuvre cet algorithme en sens inverse :



car cela montre alors une construction de la suite de Thue n'utilisant que les deux premières propriétés :

$$t_{2k} = t_k \ ; \ t_{2k+1} = 1 - t_{2k}.$$

Et on remarque que ces deux propriétés suffisent à définir la suite (avec une initialisation, par : $t_0 = 0$).

Le *deuxième algorithme* est calqué sur la construction par blocs de taille exponentiellement croissante, en considérant la suite comme une chaîne de caractères booléens (ne prenant que deux valeurs, vrai ou faux), qu'on peut traduire au bout du calcul par 0 et 1 pour retourner à la suite t. Il faut au préalable fixer le nombre n, et la construction proprement dite ne prend qu'une ligne d'instructions. Après n concaténations successives, l'algorithme donne dans l'ordre naturel les 2^n premiers termes de la suite, où n a été fixé à l'avance.

ch est une chaîne de booléens : (abcd ..z)

fonction change (ch) : (abcd ..z) ← (non a, non b, non c, non d, … non z)

m ← (vrai)

pour i de 1 à n, m ← (concaténation (m , change(m))